World Patent Information 37 (2014) 3–13

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

Review

A literature review on the state-of-the-art in patent analysis

Assad Abbas*, Limin Zhang, Samee U. Khan

North Dakota State University, USA

Keywords:
Patent analysis
Text mining
Visualization techniques
Natural language processing

A B S T R A C T

The rapid growth of patent documents has called for the development of sophisticated patent analysis tools. Currently, there are various tools that are being utilized by organizations for analyzing patents. These tools are capable of performing wide range of tasks, such as analyzing and forecasting future technological trends, conducting strategic technology planning, detecting patent infringement, determining patents quality and the most promising patents, and identifying technological hotspots and patent vacuums. This literature review presents the state-of-the-art in patent analysis and also presents taxonomy of patent analysis techniques. Moreover, the key features and weaknesses of the discussed tools and techniques are presented and several directions for future research are highlighted. The literature review will be helpful for the researchers in finding the latest research efforts pertaining to the patent analysis in a unified form.

1. Introduction

Contemporary advances in the technological arena have elevated the need for managing organizational knowledge scattered across diverse sources of information. A key challenge for knowledge management systems is the effective discovery and utilization of the contents stored in knowledge bases. Applying the same analogy to patents, it can be inferred that analyzing patents is essentially worthwhile to manage the complexities of searching and inter-relating patent information [1]. A patent represents an invention in a particular field of technology and also previous studies portray that a considerable part of the information presented in patents is relatively new [1,2].

With the ever-increasing volumes of patent information, the tasks of patent search and analysis have become vital from both legal and managerial perspectives [3]. Consequently, the patent data is analyzed in a variety of ways to fulfill different purposes. For instance, organizations are interested in analyzing patents for: (a) determining novelty in patents, (b) analyzing patent trends, (c) forecasting technological developments in a particular domain, (d) strategic technology planning, (e) extracting the information from patents for identifying the infringements, (f) determining patents quality analysis for R&D tasks, (g) identifying the promising patents, (h) technological road mapping, (i) identification of technological vacuums and hotspots, and (j) identifying technological competitors.

Various tools and techniques have been developed to assist patent analysis experts, business managers, and technology offices to fulfill diverse requirements. The task of analyzing the patent data using the automated tools to discover the patent intelligence through visualization, citation analysis, and other techniques, such as text mining is termed as patent informatics [54,55]. The techniques can be mainly classified into text mining techniques and visualization techniques. The text mining techniques further utilize Natural Language Processing (NLP) based approaches, semantic analysis based approaches, rules based approaches, property–function based approaches, and neural networks based approaches. On the other hand visualization techniques for patent analysis also use certain text mining approaches to present the results of patent analysis in visual form. The visual output of the task of patent analysis is in the form of patent networks, patent maps, and data clusters that emerge as a consequence of applying a particular algorithm. The patent map is a tool that is used to visualize the relationships among the patents by constructing the maps through the keywords and the key phrases [18]. The notion of patent networks is analogous to the widely used concept of networks. However, in patent networks the nodes represent the patents whereas the links in the network represent the relationship among the nodes or patents [8]. Clustering is a data mining concept that is used to group data items into clusters or groups based on their categories. The clustering techniques use unsupervised classification of data [59] and also have been used in patent analysis for clustering patent data according to the relevance.

* Corresponding author.
E-mail addresses: assad.abbas@my.ndsu.edu (A. Abbas), limin.zhang@ndsu.edu (L. Zhang), samee.khan@ndsu.edu (S.U. Khan).

0172-2190/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wpi.2013.12.006

We present a literature review on research efforts carried out for analyzing patents belonging to different technological areas. We also present taxonomy of techniques used to extract and analyze patent data. A few prior studies have surveyed the literature related to patent analysis. For example, the survey presented in by Bonino et al. [1] offer an insight to the contemporary patent informatics works. However, the study mainly focuses on semantic based techniques used for patent analysis. Another piece of research conducted by Saad and Nürnberger [4] studied prior-art cross lingual information retrieval approaches from patents. Liu [60] encompassed the visualization analysis of the patents and papers pertaining to the terahertz technology. Hanbury et al. [61] presented a survey that focuses on analysis and retrieval of the images, drawings, diagrams, charts, and plots from patent documents. In this literature review, we present a detailed review of major patent analysis techniques developed in recent years and classify them into text mining and visualization techniques. We also present taxonomy of the patent analysis techniques. Moreover, the strengths and weakness of the patent analysis techniques are presented and some possible future research directions are also highlighted.

The literature review is organized as follows. Section 2 presents the methodology adopted to conduct the literature review. Section 3 discusses the background and significance of patent analysis. Section 4 discusses the patent analysis techniques, and Section 5 concludes the literature review and highlights some future research directions.

2. Methodology

To carry out the literature review on patent analysis, the articles on patent analysis were searched through web searches. To find out the most relevant research work on patent analysis, the search was applied to databases, such as ScienceDirect, ACM digital library, IEEE digital library, and CiteSeerX. The search was narrowed using various terms, such as “patent analysis tools and techniques,” “visualization approaches in patent analysis,” “text mining and patent analysis.” The articles published in more recent years were selected for presenting a detailed discussion on tools and techniques used for patent analysis. The reference lists of the published research articles were also scanned. Out of the retrieved articles from the databases and based on the specified search criteria, a total of 22 articles were selected for detailed discussion in the literature review. Research articles published in last five years with a focus on the development of tools, techniques, and algorithms for analyzing patents using text mining and visualization techniques were selected. The purpose of presenting the research articles in detail is to provide the readers with the latest research on patent analysis in a unified form.

3. Background and significance of patent analysis

The existing patent search corpus comprises millions of patents scattered across different databases integrated through up-to-date web sources. For example, the most popular patent repositories for the patent documents are the United States Patent and Trademark Office (USPTO) [5], the European Patent Office (EPO) [6], and the Japan Patent Office (JPO) [7]. However, the ever increasing volumes of technical data pertaining to the inventions in certain fields of technology are difficult to analyze for the various purposes stated in the previous section. Therefore, one can no longer completely rely on experts’ knowledge and skills to analyze the patents [9] and this has consequently necessitated the use of computer aided tools for analyzing the patents [8]. Utilizing automated tools for patent analysis not only relieves the patent analysis experts of the laborious tasks of manually analyzing the patents but also speeds up the analysis process. Patent analysis involves a series of steps, including extracting patents from patent databases, extracting the information from the patents, and analyzing the extracted information to infer the logical conclusions. Fig. 1 shows the generic workflow of patent analysis. The patents contain various types of content, such as structured and unstructured data. The unstructured patent data comprises narrative text including the patent title, abstract, claims, and description. The structured patent data contains information, such as the inventor of the patent, assignee of the patent, and citation information [3,10].

However, patent analysis also entails some challenges related to the capabilities of the patent experts; whereas, the others pertain to the information presented in patents and the patent databases. Patent analysts with different levels of expertise require patent analysis tools with versatile capabilities [1]. On the other hand, the issues related to patent information are more complex and critical from the perspectives of searching the patent databases and retrieving information. The task of searching the patent databases to find relevant patents is supported by various data and text mining tools. Text mining tools with capabilities of mining text from structured and unstructured data have been developed. Mining the structured information from patent documents is relatively easier as compared to unstructured data because of its textual nature. Therefore, the task of parsing the unstructured data requires extraction tools having capabilities of segmentation of textual data into meaningful structures [10]. The visual output of the structured patent data is represented in the form of graphs and networks whereas the results from the unstructured patent data are represented as patent maps.

Patents are representatives of the technological innovations of a country or an organization and are indeed an agreement between the inventor of the patent and government or any agency designated by the government [11,12]. Therefore, analyzing the patents within a particular domain supports organizations in various aspects. For example, organizations or individuals interested in filing patents are often concerned whether a certain invention is indeed novel [1]. Patent analysis is beneficial for organizations in determining the novelty of their inventions, as well as identifying the Intellectual Property (IP) and technological competitiveness (strengths and weaknesses) of the competitors [13]. Besides the technological competitiveness, using IP information also helps in estimating the developments of a particular firm in a specific time interval [11]. Moreover, research in patent analysis has been conducted to determine the relationship between the technological advances and economic development. Furthermore, patent analysis is also useful in identifying the future technological trends in a specific field of technology.

In a patent that belongs to particular field of technology a substantial part of the information is relatively new that witnesses the novelty of the patent claims. On the other hand, organizations should investigate the potential infringement risks before investing in new products because patent litigation may possibly result in huge financial bearings. However, analyzing patents manually by domain experts to identify the infringements is labor intensive and time consuming for huge volumes of textual data [14]. Therefore, the process requires well defined automated procedures and tools to determine the infringements. There are two types of approaches for the patent infringement analysis. One approach works with patent citations, thereby looking at the patents cited by a target patent [15]. The other approach involves identifying similarities among the patent documents and converting the unstructured patent text into structured patent text using particular text mining techniques [14].

Several tools have been developed to assist business managers in identifying technological trends and formulating the strategies for

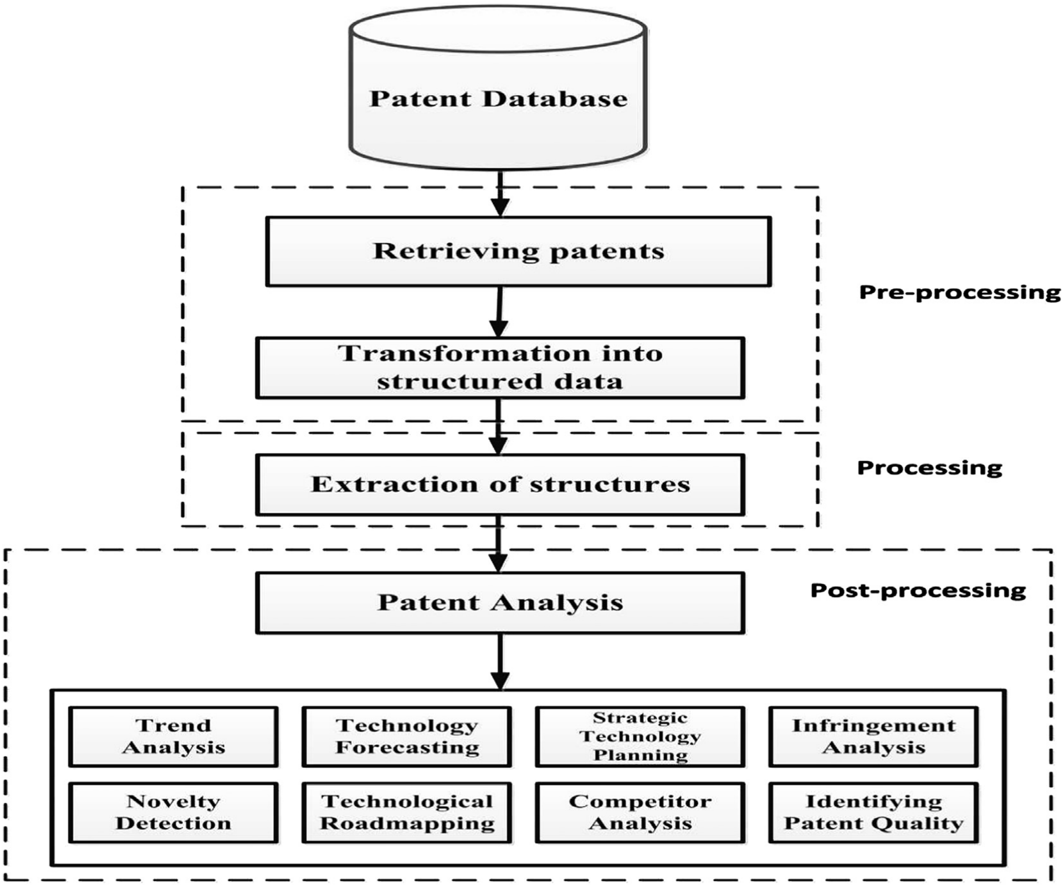

Fig. 1. Generic patent analysis workflow.

product development. The said tools are capable of generating patent maps and patent networks [17,18]. Analyzing patents through maps and networks not only provides an insight into the current technological trends but also enables the R&D policy makers to forecast future technologies and identify potential competitors. Another useful application of patent analysis tools to support R&D tasks for strategic technology planning is called Technology Roadmapping (TRM) [19]. The TRM is a methodology that is used to support the strategic research and development tasks of an organization with the aim of mapping the technological developments with the product evolution and market opportunities [19].

Moreover, patent analysis tools can free patent experts from the laborious tasks of analyzing the patent documents manually and determining the quality of patents. The tools assist organizations in making decisions of whether or not to invest in manufacturing of the new products by analyzing the quality of the filed patents [20]. Patents are also analyzed to identify the promising patents for technology transfer. The concepts of technology transfer and technology commercialization enable organizations to transfer the skills, technologies, and manufacturing techniques to the organizations having sufficient technological competence [21]. Moreover, the purpose of technology transfer is to ensure that the latest developed technologies are widely disseminated to a range of R & D communities. To transfer technology successfully, one of the key tasks is the identification of the high value technology. The methods and techniques for technology transfer that follow the TRIZ trends and text mining approaches have been proposed in Ref. [22]. The TRIZ is a Russian acronym that stands for the "Theory of Inventive Problem Solving." For problem solving and technological analysis, the TRIZ is considered as an important theory that encompasses tools, methodologies, and knowledge bases [22,56,58]. Moreover, the TRIZ trends express the evolutionary status by identifying different trend phases and also predict the improvements by analyzing and categorizing the patents [57]. In the following section we present the patent analysis techniques. Fig. 2 presents taxonomy of techniques for patent analysis.

4. Patent analysis techniques

As patents contain huge volumes of structured and unstructured data, it requires tools that are intelligent enough to accomplish the analysis tasks. There is no apparent classification of conventionally utilized tools and techniques for patent analysis. However, the vast body of literature on patent analysis has used text mining and visualization based approaches for analyzing the patent content.

The text mining techniques are used to extract the information from structured or unstructured text. The visualization techniques are meant to assist the decision makers or technology experts in representing the patent information visually to analyze the results. Therefore, we classify patent analysis techniques into text mining techniques and visualization based techniques. We also present taxonomy of the techniques for patent analysis. Table 1 presents a quick review of each piece of research work presented on patent analysis.

4.1. Text mining techniques

Text mining is a knowledge based process that uses analytical tools to derive meaningful information from the natural language text. The information is derived from the text by identifying and

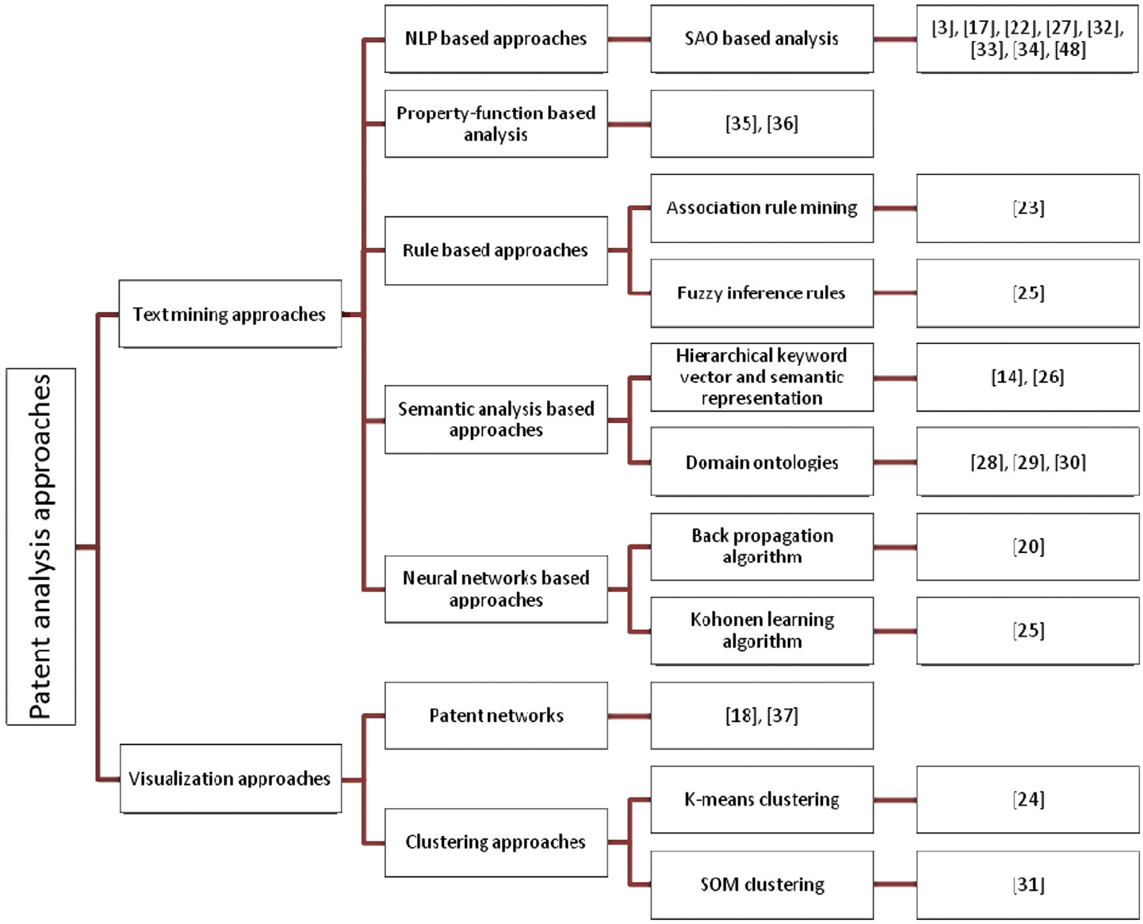

Fig. 2. Taxonomy of techniques for patent analysis.

detecting significant patterns from unknown textual data [10,38]. More specifically, text mining deals with extracting useful patterns from unstructured data rather than structured data. Text mining is used to support the processes of: (a) enterprise business intelligence for discovering the competitors [39], (b) monitoring security by analysis the text from online sources, such as internet news and blogs [40], (c) sentiment analysis [41], and (d) mining information for biomedical applications [42]. However, the text mining techniques come across the issue of correct representation of concepts that may lead to extraction of inaccurate structures from the text. Moreover, text mining approaches are also limited in correctly classifying synonyms from large documents containing the unstructured text [16]. The text mining techniques used in patent analysis are largely based on NLP, property–function based approaches, rule based approaches, neural networks based approaches, and semantic based approaches.

4.1.1. *Natural language processing (NLP) based techniques*

The NLP is a text mining approach that uses computational mechanisms to analyze and represent the textual information present in electronic documents. In patent analysis, the NLP has also been used for the transformation of the technological information into simple language structures by extracting the grammatical structures from the textual data and creating the structural relationships among the components [52]. The NLP based text mining approaches are broadly categorized into: (a) keyword based approaches and (b) Subject–Action–Object (SAO) based approaches [22]. Although keyword based text mining approaches are simple to implement, they lack in representation of important technological concepts and relationships. The keyword based approach involves predefining keywords and key phrases that require expert knowledge. On the other hand, the SAO based text mining techniques are capable of analyzing unstructured information by representing the relationships among key technological components [43]. The patent documents are transformed into SAO structures and each structure consists of a Subject (S), Action (A), and Object (O). Unlike keyword based approaches that rely on keyword vectors composed of frequency of occurrences [22]. The SAO structures are extracted directly from the patent documents [33]. The SAO structures allow the representation of the concepts in the format of problem-solution and are usually based on TRIZ [34]. However, the NLP based approaches suffer from the issues of lexical and grammatical ambiguities and also lack in representing the semantic relationships among the grammatical structures. Despite their limitations, the NLP based approaches have proved extremely effective in processing large documents containing huge volumes of textual data. The keyword based and the SAO based approaches are presented below.

Liu et al. [3] developed an integrated system for retrieval and analysis of patents called Patent Retrieval and Analysis Platform (PRAP), to help companies manage patent documents more effectively. A hybrid structure for higher search accuracy is proposed that combines bibliographic coupling and text mining approaches. Text mining is used to discover patterns and trends from huge collections of unstructured documents. The major components of PRAP are the field matching engine and the text mining engine. The field

Table 1
Comparison of patent analysis techniques.

Work	Purpose	Technique(s)	Applicable to structured/ un-structured data	Key feature(s)
Text mining approaches				
[3]	To enhance search accuracy and similarity identification	Text mining, bibliographic coupling	Structured/unstructured	Combines bibliographic coupling and text mining approaches
[17]	Identification of technological vacuums, technological hotspots and trends	NLP, patent maps, SAO structures	Structured/unstructured	Construction of dynamic patent maps to extract relationships from structured and unstructured text
[22]	Identification of promising patents	SAO based text mining	Unstructured	Capability to deal with big patent data and analyze them automatically
[27]	To extract the information particular to functionality of the patent	SAO based structures, NLP	Unstructured	Construction of patent maps and patent networks to create dependence relationships
[32]	To develop a Technology Tree (TechTree) for technology planning	Text mining, NLP	Unstructured	To develop a Technology Tree (TechTree) by mining and examining patent information
[33]	Infringement detection	SAO based similarity detection, clustering, patent map generation	Unstructured	Detecting the similarities on the basis of clustered maps
[34]	Technology road-mapping	SAO based TRM, text mining,	Unstructured	Uses grammatical structures to measure technological developments
[48]	Novelty detection	SAO based similarity detection	Unstructured	Constructs a similarity matrix to determine novelty
[35]	Identification of technological trends	Property–function based approach	Unstructured	Does not require predefined set of keywords and key phrase patterns
[36]	Identifying technological significance of new filed patents and estimating the technology pace	NLP, patent network generation, property function and co-occurrences	Unstructured	Does not require predefined set of keywords and key phrase patterns, construction of patent networks
[23]	Patent trend change mining	Association rule mining	Structured/Unstructured	Capability of mining changes in patent trends through metadata analysis
[25]	To assist in strategy planning	Kohonen learning algorithm and first nearest neighbor heuristic	Structured/unstructured	Context sensitivity
[14]	Patent infringement analysis	Hierarchical keyword vectors, tree matching algorithms	Structured/unstructured	To create dependence relationships among the claim elements
[26]	To deal with the issues of larger text and rich contents of the patent documents	Semantic analysis and text mining techniques, Naïve Bayesian Algorithm	Unstructured	Ability to extract the key concepts and identify relationships
[28]	To overcome the heterogeneity and managing information from multiple domains	Ontology for information retrieval	Unstructured	Construction of patent maps and patent networks based on semantic similarities between the patents
[29]	To deal with issues of terminological variations such as synonyms and polysemy	Information extraction using ontology	Unstructured	Capability to define semantics expressed in information silos
[30]	To classify patents according to relevance	K-nearest neighbor extraction, Ontology	Unstructured	Network based classification of patents
[20]	Determining patent quality for R&D operations	Back-propagation algorithm	Structured	Minimizes the time to determine the patent quality specific to a technology domain
Visualization approaches				
[18]	Technological trends identification	Patent Networks, bibliometric patent analysis	Structured	Uses graphs and quantitative technique for constructing networks
[37]	Important for identifying the competitors	Patent network generation	Structured	Patent ranking
[24]	To determine the trend shift for ubiquitous technologies	Clustering using K-means algorithm	Structured/unstructured	Makes a semantic network of keywords to determine meaningful relationships
[31]	To identify new research directions	Self-organizing maps	Structured	Extraction of patent terms, context retrieval, and context ranking

matching engine makes use of a bibliographic pattern discovering algorithm to discover clusters of related patent records in a collection. The text mining engine of PRAP is implemented through the vector space model. The text mining engine uses pipelines, such as Title Pipeline, Abstract Pipeline, Patent Claim Pipeline and Detail Description Pipeline. As the definition of similarity can be different for different categories of searchers, the PRAP allows users to select which pipeline is to be enabled in text mining analysis. The results from the field matching engine and text mining engine are combined through weighting model. The final result is obtained by calculating the Confidence Index (CI) of each patent record and the final score of similarity is calculated using some weighted parameters provided by the person conducting the search.

Yoon et al. [17] introduced a method to construct patent maps dynamically by analyzing the SAO based contents to identify the technological competition trends. By applying the rules of NLP, the method extracts the SAO structures and generates the patent maps. The proposed approach comprise of four steps. Patent data is collected from patent database in first step, followed by the syntactic analysis of patent documents through the NLP. The output of the syntactic analysis is presented in the form of extracted SAO structures. Consequently, the semantic similarity is measured on the basis of SAO based semantic similarities and a patent similarity matrix is constructed. Afterwards, the similarities in the SAO structures are explored through Multidimensional Scaling (MDS) and the output is visualized in the form of a dynamic patent map. The MDS refers to statistical techniques used for information visualization to discover the similarities and dissimilarities in the data. The maps generated are effective in identifying technological vacuums and technological hotspots. However, the downside of the proposed method is that the use of MDS for visualizing patent maps and *k-means* for clustering can result in information loss that eventually results in inaccurate clustering. In patent analysis, k-means is used to present visual output, such as clustering the extracted language structures to form patent maps. The *k-means* is a clustering algorithm used for assigning objects on the basis of attributes into *k* groups.

Park et al. [22] proposed a new approach to identify the potential patents for transferring technology. The TRIZ evolution trends are used for evaluation of technologies in patents. Moreover, to automatically analyze the vast patent data, the SAO based text mining technique is used. The proposed research comprises collecting patents to be analyzed from patent databases, identifying the technological life cycle stage, and extraction of the structures based on the SAO. The SAO text mining approach makes use of the NLP techniques to extract the language structures from the patent documents. On the basis of semantic similarity between the SAO structures, the evolution trends are identified. Moreover, the method is capable of identifying the promising patents. A patent is considered as a high future-value patent if that is relevant to future-important TRIZ trends. The patents are subsequently ranked based on the similarity scores and are classified. However, the approach exhibits weaknesses when making a generic classification of the TRIZ trends that eventually may not be applicable to all the technological domains. Therefore, to make the approach more effective, revision of the classification by the domain experts having knowledge of the TRIZ trends is required.

Park et al. [27] proposed SAO based intelligent patent analysis system called TechPerceptor that makes use of grammatical analysis with the NLP to extract the function information from each patent. The proposed system constructs the patent maps and patent networks on the basis of semantic analysis of patent SAO structures. The proposed system constructs patent maps and patent networks using the semantic similarities between the patents. The architecture of TechPerceptor comprises mainly three modules: (a) Patent SAO Mining Module (PSMM), (b) Patent Map based Intelligence Module (PMIM), and (c) Patent Network based Intelligent Module (PNIM). The PSMM module interacts retrieves the process data, extracts the SAO structures from patent text using NLP. Subsequently, the PSMM merges the synonyms into representative phrases and measures the semantic similarity using the ontology database for both general and domain specific terms. The PMIM constructs the patent maps based on patent similarities and identifies the technological hotspots and patent vacuums. The novelty in patents is determined by generating the patent maps and measuring the technological distances among patents. The patents located relatively far from the rest of patents are considered as outlier technology. The PNIM constructs networks using patent similarities and SAO structures to analyze the rapidly evolving technological trends. The PNIM uses Patent Network Generator (PNG) to visualize the technological relationship information in a particular technology domain.

Choi et al. [32] proposed an SAO based approach for text mining that develops a Technology Tree (TechTree) by mining and examining patent information. The information extracted through SAO structures is categorized on the basis of similarities. The two important processes of the proposed approach are the development of procedures to construct source data from patents and a method to build TechTree from that data. The NLP is used to extract the SAO structures and text mining techniques are used for analysis of SAO structures. The similarities among the SAO structures are calculated and a similarity matrix is produced. The proposed SAO-based text mining approach consists of the steps, such as (a) patent document selection, (b) Extraction of SAO structures, (c) determining the SAO similarity, (d) word phrase and action-object categorization, (e) identification of word phrase types, and (f) identification of object–action combinations. Moreover, the authors developed a TechTree analyzer to carry out the analysis using technology meta-information, such as assignee and filing date. The proposed approach was applied to generate the TechTree for Proton Exchange Fuel Cell Technology and is expected to be highly beneficial for R&D policy makers in technology planning process. However, in the proposed approach the tools to extract the SAO structures to generate patent maps are not integrated. Therefore, the task of converting the unstructured data into structured data becomes inconvenient.

Another approach to identify the infringements in patent documents on the basis of SAO structures is presented by Park et al. [33]. The authors claim that the proposed approach overcomes the inadequacies of keyword-based technological similarity determining approaches. The keyword vector based approach is limited in reflecting the particular technological key findings and the relationships between the technology components. Therefore, Park et al. [33] used SAO structures to express the structural relationships that exist between the technological components and to identify the infringements. The proposed SAO based approach collects the patent sets through NLP followed by calculating semantic similarities using WordNet. Subsequently, the patent maps are generated using the Multidimensional Scaling (MDS) on a 2-dimensional space. Moreover, a clustering algorithm is used that automatically suggests the possible infringement on a patent map. The usefulness of the proposed approach was verified by applying it to a patent infringement case in the treatment technology for prostate cancer. Although the approach uses WordNet to measure the semantic similarity in the patent documents, the WordNet database does not contain all the domain-specific terms. Therefore, there are possibilities that the approach may not completely serve the purpose of determining the patent infringements.

An approach for Technology Road-Mapping (TRM) to correlate technologies with the strategic business objectives of organizations is presented by Sungchul et al. [34]. The companies and governments that employ TRM aim to minimize the costs while sustaining objectivity. Keyword based quantitative approaches to generate TRM have been suggested but they fail in representing how technologies are employed in the relevant area and what effects they can have on other technologies. Therefore, the SAO-TRM based approach presented by Sungchul et al. [34] seems adequate for quantitative TRM. The SAO structures from patent documents are extracted through text mining and NLP based techniques. Moreover, the SAO-TRM proposed introduces the notion of "function" for supporting the quantitative patent analysis using TRM. Information on how the technology is used can be easily provided by a function. The SAO-TRM uses Product-Function-Technology (PFT) maps to assist in decision making. The purpose of PFT maps is to outline the association between the product and the direction of technological development. The validity of the approach was evaluated through a case study in proton exchange fuel cells technology. The results demonstrated that the approach is helpful for R&D managers in decision making during the TRM. However, the approach manually analyzes the technological word phrases for identification of their types. Therefore, the approach cannot completely support the process of technological road mapping based on the SAO-TRM.

Gerken and Moehrle [48] presented semantic analysis based approach to identify the inventions in patents that are highly novel. The first step in the proposed approach extracts the semantic structures from the textual patent data. The semantic structures are extracted through the syntactic analysis of the patents by using part-of-speech tagging. The authors in Ref. [48] used *Knowledgist* software [51] to extract the SAO structures. In the second step, the semantic structures are identified and linguistic analysis particular to domain and situation related elements are performed. The analysis is important to resolve the issues of synonyms that may arise from the specific domain or situation. In the third step, the similarity measure is performed to determine the novelty of the patent. A similarity matrix is created based on the comparison of the semantic structures. Once the similarity matrix has been constructed the novelty of the patent is determined by comparing the values of the matrix. The approach was evaluated through a case study in the automotive industry and the authors claimed that the semantic patent analysis is highly effective in identifying the highly novel patents. The downside of the approach is that it determines the novelty by calculating the distance between a new patent and the existing patents. However, there can be many other aspects of the similarity among patents that might be overlooked by the approach presented by the authors in Ref. [48]. Therefore, the approach needs further improvements for handling the complex relationships required to determine the novelty among patents.

4.1.2. Property–function based techniques

The property–function analysis approach extracts properties and functions from patent documents as innovation concepts through grammatical analysis. A property expresses a specific characteristic of a system whereas a function represents a suitable action of the system [47]. Unlike keyword approaches, property–function based approaches do not require predefined set of keyword and key phrase patterns. Despite their effectiveness, property–function based techniques also have issues similar to other text mining and NLP based techniques. Patent analysis systems based on property–function techniques are presented below.

Yoon and Kim [35] presented a system called TrendPerceptor for identifying the technological trends from patents. TrendPerceptor makes use of a property-function based methodology for assisting experts in identifying the inventive concepts and performing the evolution trend analysis for technology forecasting. The invention concepts are extracted through properties and functions. The properties and functions of the system are attained through grammatical analysis of the textual data. To automatically retrieve the properties and functions, the TrendPerceptor uses NLP. To facilitate the experts in analyzing technological trends, the TrendPerceptor creates the network of properties and functions. Likewise, the system assists by suggesting the improvements through automation of TRIZ trend analysis. The results from the TRIZ trend analysis depict the evolution specific to a technology that eventually helps in forecasting the technology future.

Yoon and Kim [36] proposed a Property–Function based Patent Network (PFPN) to gain understanding about the technological trends and developing the future strategies. The benefit of the property–function approach is that it eliminates the need to predefine the keywords or patterns for key phrases. The properties and functions can be mined from patent documents through natural language processing. The authors represented each patent document as a matrix that codifies properties, functions, and co-occurrences. Subsequently, the patent network is constructed by measuring similarities among patents. The network depicts the relationships that exist among patents in a given set. The approach is beneficial in identifying technological connotations, such as patents technological significance, competence of applicant filing the new patent, and the pace of technological developments of novel patents. The method was applied in silicon based thin film solar cells and the results were found encouraging in meeting the research objectives. Nevertheless, the approach exhibits certain limitations when it is used to determine the technological importance of new patents having different technological foundations.

4.1.3. Rule based techniques

Rule based techniques for text mining mostly use some sort of inference rules and association rules. Such kinds of techniques are effective for creating meaningful associations among the structures extracted from large data sets [44]. The rules are usually IF-THEN rules that help in extracting the appropriate data from the patents. However, the rule based approaches have certain issues associated, for instance rules exhibit incompetency in representing the incomplete knowledge. Moreover, as the number of the rules in the rule base increase, the risks of obtaining spurious associations among the rules also increase [53]. The systems based on rule based approaches are discussed below.

Shih et al. [23] proposed a technique called Patent Trend Change Mining (PTCM) to capture changes in patent trends without the need of specialist knowledge. The proposed approach is capable of mining changes in patent trends through metadata analysis and the changes are ranked through a degree of change. The PTCM approach consists of components, such as: (a) patent fetcher, (b) patent transformer, (c) patent indicator calculator, and (d) change detection module. Patent fetcher module uses a keyword search strategy to retrieve patents for analysis and in return gets information of assignee and International Patent Classification Code (IPC). The patent transformer module transforms the raw patent documents from HTML format into text format, stores that in database and filters out irrelevant information, such as Patent No., IPC, Application Date, Assignee Name, and Assignee Country. The patent indicator calculator module determines the patent values. The authors used four patent indicators including: (a) citation index, (b) originality, (c) generality, and (d) technology cycle time. The change detection module is the key module to determine the patent change trends. The authors used association rule mining to identify patent trends. Thus, the frequently mined patterns can be regarded as the trends extracted from patent documents. The rule matching method used computes the similarity measures and difference measures of the patent trends for two rules between two different times. In addition, the authors evaluated the degree of change and ranked the changed rules according to their importance.

Yu and Lo [25] presented a strategy planning method that integrates the patent analysis techniques with IF-THEN rules based Fuzzy Inference System (FIS). The proposed approach is context sensitive and obtains knowledge from a global patent database instead of domain experts. The important input attributes of the FIS, such as (a) Patent Quantity (PQ), (b) Revealed Patent Advantage (RPA), (c) Patent Activity (PA), (d) Be Cited Rate (BCA), and (e) Relative Citation Index (RCI) are acquired from computer aided patent analysis system. The fuzzy IF-THEN rules were used to refine the strategic rules to better fit in the construction domain. The associated parameters to infer about the appropriate technology are automatically constructed using Kohonen learning algorithm [45] and first nearest neighbor heuristic [46]. To illustrate the applicability of the proposed strategy, a case study was conducted in two different scenarios for two construction companies. The first scenario was for strategy planning of a construction firm in the domestic market whereas, the second scenario was developed for the strategy planning of the firm that was planning to enter a foreign market. The proposed method was observed suitable for planning the technological strategies with partial or incomplete information. On the other hand, the presented fuzzy inference based approach may have limited generalizing capabilities that may require expert knowledge to determine the rule set to make the system work properly.

4.1.4. *Semantic analysis based techniques*

Semantic-based text mining techniques rely on domain knowledge and create relationships among domain specific concepts [1]. The types of the techniques are effective in identifying the similarities among patents and determining the future technological trends by logically relating parsed grammatical structures. However, semantic based approaches also face problems particular to parsing the structures of natural language. Therefore, semantic analysis based approaches may exhibit incompetence in accurately representing the concepts. Various semantic based systems have been developed for patent analysis and a few are discussed below.

Lee et al. [14] proposed an approach for semantic analysis of the claims made in patent documents to identify the infringement, if any. The sections containing claims in the patents consist of semi-structured data that in reality is difficult to analyze from the perspective of infringement detection. Therefore, Lee et al. [14] emphasized on capturing the dependency relationships that occur among the elements of claims section. To represent the dependence relationships among the structured claim elements and unstructured text data, the proposed approach uses hierarchical keyword vectors. The hierarchical keyword vector utilizes similarity indicators to identify the relationship among the claim elements. Moreover, a tree matching algorithm is used that compares the elements on claim-by-claim basis. Contrary to the previous approaches that focused on comparison of technical keywords, the proposed approach effectively deals with semantics. The proposed method was validated through a case study in the DNA chip technology domain by comparing with a conventional vector based approach. Although the results provided by the proposed approach were found considerably accurate, the authors still believe that the approach is difficult to apply in other fields of technology because they may have different patenting behaviors.

Wang and Cheung [26] developed a Semantic Intellectual Property Management System (SIPMS) to cope with the problems of growths in patent documents, lengthy text, and richer contents in technological terminologies. The SIPMS has capabilities of semantic analysis and uses text mining techniques to process and analyze the patent documents. The SIPMS extracts the key concepts of the patent documents and discovers the relations among those concepts on the basis of syntactic structure of the document. The SIPMS comprises of mainly three processes, such as (a) preprocessing, (b) patent analysis, and (c) invention support. The preprocessing consists of indexing agent, segmentation agent, and indexing agent. The preprocessing process selects the relevant patents, divides patents into sections, and indexes the documents for further analysis. The extraction agent checks the relevant patent databases and if a new patent is found, it extracts the patents based on a predefined schema. The task of the segmentation agent is to divide the selected patent in a semi-structured format based on filing date, assignee, IPC codes, titles, abstracts, claims, and description of invention. Moreover, the semi-structured format is converted into concepts by an indexing agent. The patent analysis process consists of a classification agent and a relationship agent to create patent maps. The patent classification agent uses a naive Bayesian algorithm to form categories. Naïve Bayes algorithm is the simple classification method based on the Bayes rule. The Naïve Bayes algorithm makes the assumption that the particular attributes present in a class are independent of the presence of other attributes. The task of the relationship agent is to create relationships among the indexed patent documents. The third important process is invention support that is managed by a query agent and a retrieval agent. The authors used the abstracts of patent documents collected from the USTPO database to conduct experiments with the SIPMS. The experimental results depict that SIPMS is highly effective in retrieval, automatic classification, and sharing correct knowledge from massive unstructured text. However, the naïve Bayesian classifier used in the proposed approach lacks in modeling the dependencies completely that may result in inappropriate patent classification.

Taduri et al. [28] developed ontology to overcome the heterogeneity and management of the information from multiple domains, such as patent documents, court cases, and file wrapper. Through a use case in the bio domain, the authors demonstrated how the proposed ontology can be helpful for users in gathering information from multiple domains. The proposed ontology defines the semantics expressed in the information silos and serves as an integration platform. Moreover, the authors proposed to develop the knowledge base by populating the ontology classes within the information domain by appropriately relating the classes. The focus has been kept on patents issued in the U.S. The patents are available in HTML format at USTPO site. The relevant information is automatically parsed out through a script. Moreover, to download the court cases, a script is used to extract the information fields, such as plaintiff, defendant, and the court. The ontology is encoded in OWL and Protégé 3.4× is used as an ontology editor. To query the knowledge base, the RDF query language (SPARQL) was used. To evaluate the data extraction, a random sample of 50 patents was generated and parsed using the parser. The results were found to be extremely accurate in extraction of patent data from three different domains. Nonetheless, the approach has certain restrictions and therefore may not perform well. For example, the approach uses the parser that automatically extracts the patent data. Since all the patent data is not in a unified format, the task of automatic extraction of data becomes more challenging.

Taduri et al. [29] proposed a knowledge based software framework to facilitate the retrieval of patent related information from multiple, diverse, and un-coordinated information sources in the U.S patent system. To provide the interoperability among various information sources, the authors propose patent system ontology. An important issue that arises in patent analysis is of terminological variations, such as synonyms and polysemy that hinders the traditional Information Retrieval (IR) based methods. To resolve these problems, the authors proposed a knowledge based framework that uses external knowledge sources, for instance, the domain ontology to provide the required semantics. The ontology is populated from actual physical documents belonging to the document repository. Moreover, the knowledge base also contains a file wrapper including information, such as first amendment, rejection, interference, and the original application. In addition, the proposed patent system ontology applies information obtained from one domain to another. An information retrieval framework is built on top of the semantics of the patent system ontology in multiple stages that enhances multi-source information retrieval. As a result, the proposed system ontology gives a standardized representation and a shared vocabulary of information sources to facilitate interoperability.

A novel patent network based classification method to analyze patents and to predict the patent classes is presented by Shih and Liu [30]. The patent network consists of varying types of nodes to represent different features. The proposed patent classification approach is implemented in two steps: (a) patent network construction and (b) patent network analysis that includes *k-nearest* neighbor extraction and patent class identification. The patent ontology network construction step identifies the relationship between the instances/nodes. The proposed ontology network contains four types of instances/nodes and eight types of relations/edges. To classify a patent document, the algorithm determines all the connections and weights between the query patent and the nodes in the patent ontology network. The algorithm for patent network analysis calculates the weights of the nodes and their relationships to derive correlations in the metadata. Once the relevance of the query patent document to other nodes in the patent ontology has been determined, the *k-nodes* with highest relevance to the query are extracted to identify the most appropriate class of patent. The patent and class nodes are used to determine the scores of candidate classes as they are best appropriate for interpreting classes. Therefore, for a patent node, the more the relevance of the patent node to the query node, the greater the probability of that query patent of belonging to the class of that node. The experimental results of the proposed approach show that the patent network based approach is highly effective as compared to other approaches, such as content-based approaches, citation-based approaches, and metadata based approaches in terms of accuracy, precision, and recall.

4.1.5. Neural networks based technique

Neural network based approaches have also been used for patent classification and technology forecasting [50]. More specifically, the back propagation neural network algorithm has been used to train a patent network to determine the quality of patents. The neural network based approaches have also been used in conjunction with rule based approaches. Below we present one text mining approach based on artificial neural networks.

The research conducted by Trappey et al. [20] focused on minimizing the efforts and the time required to search for and to determine the patent quality to manage the R&D operations particular to an innovation. The authors extracted patent indicators from sources, such as International Patent Classification (IPC) and the number of patent citations. Moreover, the patent quality models developed along with the identification of indicators are subsequently provided as input for training through back-propagation neural networks. The purpose of training through a back-propagation algorithm is to identify the patents that are specific to a technology and to make an accurate recommendation. The patents identified are then ranked to help understand the technical worth of the patents. The analytical results of the proposed methodology were found to be 85% accurate. However, the approach may suffer from the cold start problem. The cold start problem occurs when the system initially has less data for making recommendations that eventually may result in imprecise recommendation of patents.

4.2. Visualization techniques

Another major approach for contemporary patent analysis is the use of visualization tools to represent patent information and result analysis. For instance, to understand the technological trends in a particular domain, patent maps or clustering methods can be utilized. Similarly, another visualization method called patent network is helpful in analyzing the patents to determine the similarities or infringements. Although visualization techniques visually represent the information extracted from the patents, they still use certain text mining approaches to extract the information from patent documents. However, visualization techniques using text mining approaches suffer from the similar issues as presented in the section pertaining to the text mining approaches. Below we present visualization based techniques for patent analysis.

Chang et al. [18] presented a framework to identify the technological trends by analyzing patents for Carbon Nanotube Field Emission Display (CNT-FED). To determine the present technological standing of CNT-FED, bibliometric patent analysis is performed. Afterwards, patent network analysis is performed to determine the relationships among the patents. Graphs and quantitative techniques are used for presenting information from the patent networks. The authors used the steps presented in Ref. [8] for graphing the network. In first step, the related patent keywords are selected by the experts. The frequencies of the occurrence of keywords in patent documents are calculated in second step. The relationships between the patents are established by calculating the Euclidean distance. Euclidean distance is the distance between two data points in data sets where each data point has multiple attributes. Euclidean distance is also used in data mining to accomplish the task of clustering. Moreover, the overall technological trends in the patent network are measured by using Technology Cycle Time (TCT). Technology cycle time represents the technological progress between two time intervals. In case of patent analysis TCT refers to the time interval between a previously filed patent and a target patent. A smaller value of TCT indicates a rapid growth of the technological process.. Furthermore, the comparisons amongst the clusters are calculated through density index. The analysis of clusters may be important to manage and characterize technology trends.

Contrary to the existing topic models, the authors in Ref. [37] proposed a model called Inventor–Company–Topic (ICT) model that incorporates information about the inventors and companies. The presented framework is capable of mining information from heterogeneous patent networks. The topical evolution of the objects of the patent network is modeled through a dynamic probabilistic model. Subsequently, a heterogeneous co-ranking approach is presented that ranks the various objects. The co-ranking may be helpful for new companies in entering new business market or devising the novel ideas. Another contribution of the authors is defining the measures for identifying the topic-level competitors and to identify the patterns that emerge as a result of topic modeling. This helps organizations in identifying the correct business competitors. Further, an automated summary of the search results is presented through a maximum coverage method. The maximum coverage model uses concept extraction that involves scoring on the basis of term frequencies. The experimental results show that the framework is quiet feasible for adopting in practice.

Kim et al. [24] presented a visualization method for patent analysis. The keywords from a patent document of the target technology field are collected and clustered by *k*-means algorithm.

Based on the clustering results, a semantic network of keywords is formed irrespective of patent filing dates. A map is then built up by re-arranging each keyword of the semantic network according to the earliest filing dates and frequency in patent documents. The contribution of the proposed approach is the establishment of a map that considers both structured and un-structured items of a patent document. Moreover, the formation of a semantic network of keywords from patent documents makes the proposed approach different from the previous approaches. The proposed visualization approach was tested by targeting the ubiquitous computing technology as an emerging technology. First the keywords were collected from various experts of ubiquitous computing technology and on the basis of those selected keywords, the patent documents related to ubiquitous computing technology were searched and the predefined keywords were investigated. The keywords recommended from the experts and those predefined in patent documents were merged. The existence of each keyword in the searched patent documents was determined and based on that, 96 patent documents were clustered using the *k*-means algorithm. A semantic network was built by increasing the number of groups and selected a semantic network with five groups. Based on the semantic network, the authors investigated the earliest filing date and the frequency of each node in the semantic network. Therefore, from the patent map, the trend shift of technologies related to ubiquitous computing was determined. However, the approach may suffer from the issues particular to the k-means, such as differing sizes and densities and empty clusters.

Segev and Kantola [31] developed a model to facilitate governments, inventors, and manufacturing organizations to identify new research directions. The model comprises steps such as patent knowledge extraction, knowledge representation, and identification of research trends. The key features of patent are extracted in a knowledge extraction step. The authors used the context recognition algorithm proposed in Refs. [49], which extracts the patent terms and consists of the processes, such as: (a) context retrieval, (b) context ranking, and (c) context selection. The knowledge representation step generates and evolves maps that are based on the Self-Organization Map (SOM) technique. The SOM implementation uses a vector to represent all the relevant patent descriptors with the weights that define the relevance of the patent descriptors. In the third step, the research trends are identified on the basis of knowledge extracted. Unified Distance Matrix (U-matrix) value and context descriptor are used for identification of the current patent trends. The U-matrix represents SOMs and provides insights to understand the invisible relationships in high dimensional data sets. The experimental results with the proposed model based on the SOM clustering method were more accurate as compared to *k-Means* and DBSCAN clustering methods. The DBSCAN is a density-based clustering algorithm that finds the number of clusters from the estimated density distribution of the corresponding nodes.

5. Conclusions and future research directions

With recent technological developments, patent analysis plays an ever-increasing role in defining business strategies and supporting decision making in and across organizations. Tools with versatile capabilities to effectively retrieve and visually represent patent information are beneficial to organizations in a wide range of tasks. This literature review provides an overview of text mining and visualization based techniques for patent analysis and a taxonomy that classifies these approaches.

Despite the fact that the techniques for patent analysis have become mature, there are areas that still need improvements. For instance, the tools using SAO based extraction techniques also extract certain irrelevant structures [34]. Therefore, the existing SAO based approaches need to be improved in a way that only the semantically relevant information is retrieved. Moreover, the hybrid approaches used for patent retrieval can also be utilized to search for the documents other than patents, such as journal articles [3]. Furthermore, the patent analysis approaches for strategic technology planning developed so far are capable of suggesting one strategy only. It would be worthwhile for the managers if the approaches are made more efficient and flexible to offer multiple suggestions for devising strategies [25].

Acknowledgments

The authors are thankful to Saif-ur-Rehman Malik, Ahmad Fayyaz, and Saeeda Usman for the valuable reviews, suggestions, and comments.

References

[1] Bonino D, Ciaramella A, Corno F. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. World Pat Inf 2010;32(1):30–8.
[2] Hunt D, Nguyen L, Rodgers M. Patent searching: tools and techniques. John Wiley and Sons; 2007.
[3] Liu SH, Liao HL, Pi SM, Hu JW. Development of a patent retrieval and analysis platform—a hybrid approach. Expert Syst Appl 2011;38(6):7864–8.
[4] Saad F, Nürnberger A. Overview of prior-art cross-lingual information retrieval approaches. World Pat Inf 2012:304–14.
[5] The United States Patent and Trademark Office. http://www.uspto.gov/, [accessed 28.03.13].
[6] European Patent Office. http://www.epo.org/, [accessed 28.04.13].
[7] Japan Patent Office. http://www.jpo.go.jp/, [accessed 28.03.13].
[8] Yoon B, Park Y. A text-mining-based patent network: analytical tool for high-technology trend. J High Technol Manag Res 2004;15:37–50.
[9] Kostoff R. The use and misuse of citation analysis in research evaluation. Scientometrics 1998;43(1):27–43.
[10] Tseng YH, Lin CJ, Lin YI. Text mining techniques for patent analysis. Inform Process Manag 2007;43(5):1216–47.
[11] Ernst H. Patent information for strategic technology management. World Pat Inf 2003;25(3):233–42.
[12] Hufker T, Alpert F. Patents: a managerial perspective. J Prod Brand Manag 1994;3(4):44–54.
[13] Abraham BP, Moitra SD. Innovation assessment through patent analysis. Technovation 2001;21(4):245–52.
[14] Lee C, Song B, Park Y. How to assess patent infringement risks: a semantic patent claim analysis using dependency relationships. Technol Anal Strat Manag 2013;25(1):23–38.
[15] Sternitzke C, Bartkowski A, Schramm R. Visualizing patent statistics by means of social network analysis tools. World Pat Inf 2008;30:115–31.
[16] Witten IH. Text mining. http://www.cs.waikato.ac.nz/~ihw/papers/04-IHW-Textmining.pdf, [accessed 09.10.13].
[17] Yoon J, Park H, Kim K. Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis. Scientometrics 2013:1–19.
[18] Chang PL, Wu CC, Leu HJ. Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display. Scientometrics 2010;82(1):5–19.
[19] Phaal R, Farrukh CJP, Mills JF, Probert DR. Customizing the technology roadmapping approach. Manag Eng Technol 2003:361–9.
[20] Trappey AJC, Trappey CV, Wu C—YW, Fan CY, Lin Y-L. Intelligent patent recommendation system for Innovative design collaboration. J Netw Comput Appl 2013;36:1441–50.
[21] Du R, Ai S. Cross-organizational knowledge acquisition through flexible Hiring and joint R&D: insights from a survey in China. Expert Syst Appl 2008;35:434–41.
[22] Park H, Ree JJ, Kim K. Identification of promising patents for technology transfers using TRIZ evolution trends. Expert Syst Appl 2013:736–43.
[23] Shih MJ, Liu DR, Hsu ML. Discovering competitive intelligence by mining changes in patent trends. Expert Syst Appl 2010;37(4):2882–90.
[24] Kim YG, Suh JH, Park SC. Visualization of patent analysis for emerging technology. Expert Syst Appl 2008;34(3):1804–12.
[25] Yu WD, Lo SS. Patent analysis-based fuzzy inference system for technological strategy planning. Automat Constr 2009;18(6):770–6.
[26] Wang WM, Cheung CF. A semantic-based intellectual property management system (SIPMS) for supporting patent analysis. Eng Appl Artif Intell 2011;24(8):1510–20.
[27] Park H, Kim K, Choi S, Yoon J. A patent intelligence system for strategic technology planning. Expert Syst Appl 2013:2373–90.

[28] Taduri S, Lau GT, Law KH, Yu Hang, Kesan JP. Developing an ontology for the US patent system. In: In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times 2011. pp. 157–66.
[29] Taduri S, Lau GT, Law KH, Hang Yu, Kesan JP. A patent system ontology for facilitating retrieval of patent related information. http://eil.stanford.edu/publications/sid/icegov_2012.pdf, [accessed 12.04.13].
[30] Shih MJ, Liu DR. Patent classification using ontology-based patent network analysis. In: Pacific Asia Conference on Information Systems, Taiwan 2010. pp. 962–72.
[31] Segev A, Kantola J. Identification of trends from patents using self-organizing maps. Expert Syst Appl 2012:13235–42.
[32] Choi S, Park H, Kang D, Lee JY, Kim K. An SAO-based text mining approach to building a technology tree for technology planning. Expert Syst Appl October 2012:11443–55.
[33] Park H, Yoon J, Kim K. Identifying patent infringement using SAO based semantic technological similarities. Scientometrics 2012;90(2):515–29.
[34] Sungchul C, Kim H, Yoon J, Kim K, Lee JY. An SAO-based text-mining approach for technology roadmapping using patent information. R&D Manag 2013;43(1):52–74.
[35] Yoon J, Kim K. TrendPerceptor: a property–function based technology intelligences system for identifying technology trends from patents. Expert Syst Appl 2012;39(3):2927–38.
[36] Yoon J, Kim K. An analysis of property–function based patent networks for strategic R&D planning in fast-moving industries: the case of silicon-based thin film solar cells. Expert Syst Appl 2012:7709–17.
[37] Tang J, Wang B, Yang Y, Hu P, Zhao Y, Yan X, et al. Patentminer: topic-driven patent analysis and mining. In: In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 2012. pp. 1366–74.
[38] Ghazinoory S, Ameri F, Farnoodi S. An application of the text mining approach to select technology centers of excellence. Technol Forecast Soc Change 2012;80:918–31.
[39] Coussement K, Poel DVD. Integrating the voice of customers through call center emails into a decision support system for churn prediction. Inform Manag 45(3):164–74. http://dx.doi.org/10.1016/j.im.2008.01.005.
[40] Zanasi A. Virtual weapons for real wars: text mining for national security. In: Proceedings of the International Workshop on Computational Intelligence in Security for Information Systems CISIS'08. Springer Berlin Heidelberg; 2009. pp. 53–60.
[41] Liu B, Zhang L. A survey of opinion mining and sentiment analysis. In: Mining text data 2012. pp. 415–63.
[42] Cohen KB, Hunter L. Getting started in text mining. PLoS Comput Biol 4(1): e20. http://dx.doi.org/10.1371/journal.pcbi.0040020. PMC 2217579. PMID 18225946.
[43] Yoon J, Kim K. Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. Scientometrics 2011:1–17.
[44] Han J, Kamber M. Mining association rules in large databases. In: Data mining concepts and techniques. San Francisco, CA: Morgan Kaufmann; 2001.
[45] Kohonen T. Self-organization and associative memory. Berlin,German: Springer-Verlag; 1988.
[46] Lin CT, Lee CSG. Neural-network-based fuzzy logic control and decision system. IEEE Trans Comput 1991;40(12):1320–36.
[47] Dewulf S. Directed variation of properties for new or improved function product DNA–a base for connect and develop. Procedia Eng 2011;9:646–52.
[48] Gerken JM, Moehrle MG. A new instrument for technology monitoring: novelty in patents measured by semantic patent analysis. Scientometrics 2012;91(3):645–70.
[49] Segev A, Leshno M, Zviran M. Internet as a knowledge base for medical diagnostic assistance. Expert Syst Appl 2007;33(1):251–5.
[50] Che HC, Lai YH, Wang SY. Assessment of patent legal value by regression and back-propagation neural network. Int J Syst Innovation 2009;1:31–47.
[51] Knowledgist, http://inventionmachine.com/default.aspx, [accessed 21.05.13].
[52] Masiakowski P, Wang S. Integration of software tools in patent analysis. World Pat Inf 2013;35(2):97–104.
[53] Gionis A, Mannila H, Mielikäinen T, Tsaparas P. Assessing data mining results via swap randomization. ACM Trans Knowl Discov Data (TKDD) 2007;1(3):14.
[54] Moehrle MG, Walter L, Bergmann I, Bobe S, Skrzipale S. Patinformatics as a business process: a guideline through patent research tasks and tools. World Pat Inf 2010;32(4):291–9.
[55] Trippe AJ. Patinformatics: tasks to tools. World Pat Inf 2003;25(3):211–21.
[56] Barry K, Domb E, Michael S. TRIZ – what is TRIZ. The TRIZ; http://cmapspublic.ihmc.us/rid=1206064509716_727387479_10719/TRIZ%20-%20What%20Is%20TRIZ.pdf, [accessed 29.10.13].
[57] Verhaegen P–A, D'hondt J, Vertommen J, Dewulf S, Duflou JR. Relating properties and functions from patents to TRIZ trends. CIRP J Manuf Sci Technol 2009;1(3):126–30.
[58] Yoon J, Kim K. An automated method for identifying TRIZ evolution trends from patents. Expert Syst Appl 2011;38(12):15540–8.
[59] Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv (CSUR) 1999;31(3):264–323.
[60] Liu G. Visualization of patents and papers in terahertz technology: a comparative study. Scientometrics 2013;94(3):1037–56.
[61] Hanbury A, Bhatti N, Lupu M, Mörzinger R. Patent image retrieval: a survey. In: Proceedings of the 4th ACM workshop on Patent information retrieval 2011. pp. 3–8.

Assad Abbas completed Master of Science in Informatics from University of Skovde, Sweden in 2010. Currently, he is enrolled in PhD program at Department of Electrical and Computer Engineering, North Dakota State University, USA. He is affiliated with COMSATS Institute of Information Technology, Pakistan since 2004. His research interests are mainly but not limited to Cloud Computing, Information Systems, Information Retrieval, Knowledge Engineering, and Knowledge Representation.

Limin Zhang is an assistant professor of Management Information Systems in the Accounting, Finance, and Information Systems Department at North Dakota State University. She received her PhD in Management Information Systems from the University of Arizona. Her research interests include context-based Web search, virtual team collaboration, and using technology to support competitive intelligence and decision-making. Her research work has appeared in a number of journals and various conference proceedings. Her teaching interests include database design, Web system development, and information technology management.

Samee U. Khan received a BS degree from Ghulam Ishaq Khan Institute of Engineering Sciences and Technology, Topi, Pakistan, and a PhD from the University of Texas, Arlington, TX, USA. Currently, he is Assistant Professor of Electrical and Computer Engineering at the North Dakota State University, Fargo, ND, USA. Prof. Khan's research interests include optimization, robustness, and security of: cloud, grid, cluster and big data computing, social networks, wired and wireless networks, power systems, smart grids, and optical networks. His work has appeared in over 225 publications. He is a Fellow of the Institution of Engineering and Technology (IET, formally IEE), and a Fellow of the British Computer Society (BCS).